Lecture 6

Central Limit Theorem for Yule-Walker estimators

The goal of this lecture is to prove the asymptotic normality of the Yule-Walker estimator of the coefficients of an $AR(p)$. We shall use the same notation as those of Lecture 5.

Theorem 1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a zero-mean causal autoregressive process of order p:*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$$

with $Z \sim IID(0, \sigma^2)$. Let $\widehat{\boldsymbol{\phi}}_p$ denote the Yule-Walker estimator of $\boldsymbol{\phi}_p = (\phi_1, \ldots, \phi_p)$. Then $\widehat{\boldsymbol{\phi}}_p$ is asymptotically normally distributed:

$$\sqrt{n}(\widehat{\boldsymbol{\phi}}_p - \boldsymbol{\phi}_p) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2 C_p^{-1}), \tag{1}$$

$$\tag{2}$$

where C_p is the covariance matrix of $(X_1, \ldots, X_p)$.

Nota Bene: C_p^{-1} is well defined (See Sheet 3).
In the following, we will admit these results:

(i) $\widehat{c}_n(h) \xrightarrow[n\to\infty]{\mathbb{P}} c(h)$, for all $h \geq 0$.

(ii) $\frac{1}{n} X'Y \xrightarrow[n\to\infty]{\mathbb{P}} \mathbf{c}_p$ where $Y = (X_1, \ldots, X_n)'$ and X is the matrix defined by

$$X = \begin{pmatrix} X_0 & X_{-1} & \ldots & X_{1-p} \\ X_1 & X_0 & \ldots & X_{2-p} \\ \vdots & \vdots & & \vdots \\ X_{n-1} & X_{n-2} & \ldots & X_{n-p} \end{pmatrix}.$$

A proof can be found in [1]. Here we only notice that (i) and (ii) hold true in at least two cases:

- If $X = (X_t)_{t\in\mathbb{Z}}$ is a moving average process

$$X_t = \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j}, \quad Z \sim IID(0, \sigma^2),$$

with $(\psi_j)_{j\in\mathbb{Z}}$ such that $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$ and $\mathbb{E}[Z_t^4] < \infty$.

- If $X = (X_t)_{t\in\mathbb{Z}}$ is the moving average process

$$X_t = \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j}, \quad Z \sim IID(0, \sigma^2),$$

with $(\psi_j)_{j\in\mathbb{Z}}$ such that $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$ and $\sum_{j\in\mathbb{Z}} |j||\psi_j|^2 < \infty$.

Moreover, since we will need several properties of the convergence in probability, let us quickly recall the main facts and definitions. (See e.g. [1] for the proofs.)

Definition 1. *We say that X_n converges in probability to zero, written $X_n = o_{\mathbb{P}}(1)$ or $X_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$, if for every $\varepsilon > 0$,*

$$\mathbb{P}(|X_n| > \varepsilon) \xrightarrow[n\to\infty]{} 0.$$

We say that $X_n = o_{\mathbb{P}}(a_n)$ if $X_n = a_n Y_n$ with $Y_n = o_{\mathbb{P}}(1)$.

Definition 2. *A sequence $(X_n)_{n\geq 1}$ is* bounded in probability *(or* tight*) if for every $\varepsilon > 0$ there exists $M \in (0, \infty)$ such that*

$$\sup_n \mathbb{P}(|X_n| > M) < \varepsilon.$$

We shall write $X_n = O_{\mathbb{P}}(1)$ to say that $(X_n)_{n\geq 1}$ is tight. We say that $X_n = O_{\mathbb{P}}(a_n)$ if $X_n = a_n Y_n$ with $Y_n = O_{\mathbb{P}}(1)$.

Proposition 1. *Let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be sequences of random variables defined on the same probability space and $a_n > 0$, $b_n > 0$ for all $n \geq 1$. The following properties hold*

(i) If $X_n = o_{\mathbb{P}}(a_n)$ and $Y_n = o_{\mathbb{P}}(n)$ then $X_n Y_n = o_{\mathbb{P}}(a_n b_n)$, $X_n + Y_n = o_{\mathbb{P}}(\max(a_n, b_n))$ and $|X_n|^r = o_{\mathbb{P}}(a_n^r)$, for $r > 0$.

(ii) If $X_n = o_{\mathbb{P}}(a_n)$ and $Y_n = O_{\mathbb{P}}(b_n)$, then $X_n Y_n = o_{\mathbb{P}}(a_n b_n)$.

(iii) The statement (i) remains valid if $o_{\mathbb{P}}$ is everywhere replaced by $O_{\mathbb{P}}$.

Lemma 1 (Slutsky Lemma)**.** *If the sequence $(X_n)_{n\in\mathbb{N}}$ weakly converges to X and the sequence $(Y_n)_{n\in\mathbb{N}}$ converges in probability to a constant c, then:*

- *$(X_n + Y_n)_{n\in\mathbb{N}}$ weakly converges to $X + c$;*
- *$(X_n Y_n)_{n\in\mathbb{N}}$ weakly converges to cX;*
- *$(\frac{X_n}{Y_n})_{n\in\mathbb{N}}$ weakly converges to $\frac{X}{c}$ provided that $c \neq 0$.*

Theorem 2 (Continuous mapping theorem)**.** *Suppose that $f : \mathbb{R}^k \to \mathbb{R}^m$ is almost surely continuous, i.e. f is continuous on S for some S with $\mathbb{P}(X \in S) = 1$. Then:*

- *$(X_n)_{n\in\mathbb{N}}$ weakly converges to X implies that $(f(X_n))_{n\in\mathbb{N}}$ weakly converges to $f(X)$;*
- *$(X_n)_{n\in\mathbb{N}}$ converges in probability to X implies that $(f(X_n))_{n\in\mathbb{N}}$ converges in probability to $f(X)$;*

- *$(X_n)_{n\in\mathbb{N}}$ converges a.s. to X implies that $(f(X_n))_{n\in\mathbb{N}}$ converges a.s. to $f(X)$.*

Proof of Theorem 1. To shorten the notation we will only write $\boldsymbol{\phi}$ instead of $\boldsymbol{\phi}_p$. We shall use the following notations: $Y = (X_1, \dots, X_n)'$, $Z = (Z_1, \dots, Z_n)'$ and X is the $n \times p$ design matrix, i.e.

$$X = \begin{pmatrix} X_0 & X_{-1} & \dots & X_{1-p} \\ X_1 & X_0 & \dots & X_{2-p} \\ \vdots & \vdots & & \vdots \\ X_{n-1} & X_{n-2} & \dots & X_{n-p} \end{pmatrix}.$$

The proof is divided into two main steps that imply the thesis by means of Slutsky lemma.

Step 1: we prove that the asymptotic normality holds for the "linear regression estimate" $\boldsymbol{\phi}^*$ of $\boldsymbol{\phi}$ defined as

$$\boldsymbol{\phi}^* = (X'X)^{-1}X'Y.$$

Using that $Y = X\boldsymbol{\phi} + Z$, we obtain that

$$\boldsymbol{\phi}^* - \boldsymbol{\phi} = (X'X)^{-1}X'Z. \tag{3}$$

Define $M_n^{(i)}$ as the i-th component of the vector $X'Z$, that is

$$M_n^{(i)} = X_{1-i}Z_1 + \dots + X_{n-i}Z_n, \quad i = 1, \dots, p.$$

Observe that $M_n^{(i)}$ is a martingale in n w.r.t. the filtration $\mathscr{F}_n := \sigma(Z_1, \dots, Z_n, X_0, \dots, X_{-p+1})$. Indeed:

- $\mathbb{E}[|M_n^{(i)}|] < \infty$: $\mathbb{E}[|X_{k-i}Z_k|] \le \sqrt{\mathbb{E}[X_{k-i}^2]}\sqrt{\mathbb{E}[Z_k^2]} < \infty$ for all $k = 1, \dots, n$ and for all $i = 1, \dots, p$. In fact, since by assumption X is a causal AR(p), from Lecture 3 we know that X admits the representation $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, with $\sum_{j=0}^{\infty} |\psi_j| < \infty$. From Theorem 1 in Lecture 2 it thus follows that $\mathbb{E}[X_t^2] < \infty$, $\forall t$. Since L_1 is a vector space we get that $M_n^{(i)} = X_{1-i}Z_1 + \dots + X_{n-i}Z_n \in L_1$ for all $i = 1, \dots, p$.

- $\mathbb{E}[M_n^{(i)}|\mathscr{F}_{n-1}] = M_{n-1}^{(i)}$: it is enough to notice that $\mathbb{E}[Z_n|\mathscr{F}_{n-1}] = \mathbb{E}[Z_n] = 0$. Indeed:

  $$\mathbb{E}[M_n^{(i)}|\mathscr{F}_{n-1}] = X_{1-i}Z_1 + \dots + X_{n-1-i}Z_{n-1} + X_{n-i}\mathbb{E}[Z_n|\mathscr{F}_{n-1}] = X_{1-i}Z_1 + \dots + X_{n-1-i}Z_{n-1}.$$

Let us now compute the quadratic variation of $M_n^{(i)}$, i.e.

$$\begin{aligned} \langle M^{(i)} \rangle_n &:= \sum_{k=1}^{n} \mathbb{E}\big[(M_k^{(i)} - M_{k-1}^{(i)})^2 | \mathscr{F}_{k-1}\big] \\ &= \sum_{k=1}^{n} \mathbb{E}\big[(X_{1-i}Z_1 + \dots + X_{k-i}Z_k - (X_{1-i}Z_1 + \dots + X_{k-1-i}Z_{k-1}))^2 | \mathscr{F}_{k-1}\big] \\ &= \sum_{k=1}^{n} \mathbb{E}\big[X_{k-i}^2 Z_k^2 | \mathscr{F}_{k-1}\big] = \sum_{k=1}^{n} X_{k-i}^2 \mathbb{E}\big[Z_k^2 | \mathscr{F}_{k-1}\big] = \sigma^2 \sum_{k=1}^{n} X_{k-i}^2 = \sigma^2 (X'X)_{i,i}. \end{aligned}$$

Set $M_n := (M_n^{(1)}, \dots, M_n^{(p)})' \in \mathbb{R}^p$. Then, with the same kind of computations as before, we get that the quadratic covariance matrix of M_n is given by:

$$\langle M \rangle_n := \sum_{k=1}^{n} \mathbb{E}[(M_k - M_{k-1})(M_k - M_{k-1})' | \mathscr{F}_{k-1}] = \sigma^2 X'X.$$

Therefore, using (3), we deduce that

$$\boldsymbol{\phi}^* - \boldsymbol{\phi} = \sigma^2 \langle M \rangle_n^{-1} M_n. \tag{4}$$

Recall that the goal is to prove the asymptotic normality of $\phi^* - \phi$. To that aim we shall use a vectorial form of the central limit theorem (CLT) for martingale difference series (MDS) discussed in Lecture 4. The idea is to apply the CLT for MDS to the vectors

$$\xi_j := (\sigma^2 C_p)^{-1/2}(M_j - M_{j-1}) \in \mathbb{R}^p, \quad 1 \le j \le n.$$

First of all let us observe that $(\xi_j)_{j=1}^n$ is a martingale difference scheme. Indeed, clearly ξ_j is $\mathscr{F}_j$-measurable and

$$\begin{aligned}
\mathbb{E}[\xi_j | \mathscr{F}_{j-1}] &= (\sigma^2 C_p)^{-1/2} \mathbb{E}[M_j - M_{j-1} | \mathscr{F}_{j-1}] \\
&= (\sigma^2 C_p)^{-1/2} \big(\mathbb{E}[M_j^{(1)} - M_{j-1}^{(1)} | \mathscr{F}_{j-1}], \dots, \mathbb{E}[M_j^{(p)} - M_{j-1}^{(p)} | \mathscr{F}_{j-1}] \big)' \\
&= (0, \dots, 0)
\end{aligned}$$

because, for any $i = 1, \dots, p$,

$$\mathbb{E}\big[M_j^{(i)} - M_{j-1}^{(i)} | \mathscr{F}_{j-1}\big] = \mathbb{E}[X_{j-i} Z_j | \mathscr{F}_{j-1}] = X_{j-i} \mathbb{E}[Z_j | \mathscr{F}_{j-1}] = 0.$$

Let us now show that the martingale difference series $(\xi_j)_{j=1}^n$ satisfies the hypotheses for the martingale CLT (in a vectorial form). Firstly, let us observe that the conditional covariance matrix is given by

$$\frac{1}{n}(\sigma^2 C_p)^{-1} \langle M \rangle_n = \frac{1}{n}(\sigma^2 C_p)^{-1} \sigma^2 (X'X) = C_p^{-1} \frac{1}{n} X'X.$$

Using that $\widehat{C}_p$ converges in probability to C_p and that $\frac{1}{n}X'X - \widehat{C}_p \xrightarrow{\mathbb{P}} 0$ (we shall prove that later), we get

$$\frac{1}{n} X'X = \widehat{C}_p + \frac{1}{n} X'X - \widehat{C}_p \xrightarrow{\mathbb{P}} C_p.$$

Therefore,

$$\frac{1}{n} (\sigma^2 C_p)^{-1} \langle M \rangle_n \xrightarrow{\mathbb{P}} I_p, \tag{5}$$

where $I_p = \operatorname{diag}(1, \dots, 1) \in \mathbb{R}^{p \times p}$. Let us now check the conditional Lindeberg condition, that is let us prove that for any $\varepsilon > 0$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\Big[\|\xi_i\|^2 \mathbf{1}_{\|\xi_i\| > \varepsilon\sqrt{n}} | \mathscr{F}_{i-1} \Big] \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{6}$$

Notice that to deduce (7) it is enough to prove

$$\mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\|\xi_i\|^2\mathbf{1}_{\|\xi_i\|>\varepsilon\sqrt{n}}|\mathscr{F}_{i-1}\Big]\Big] \xrightarrow[n\to\infty]{} 0, \tag{7}$$

as we shall. Indeed, using that X is a strictly stationary process (it follows from the assumption $Z \sim IID(0,\sigma^2)$), we have that

$$\|\xi_i\| = \|(\sigma^2 C_p)^{-1/2}(M_i - M_{i-1})\| \stackrel{d}{=} \|(\sigma^2 C_p)^{-1/2}M_1\|.$$

Hence, (7) follows by dominated convergence theorem (DCT):

$$\mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\|\xi_i\|^2\mathbf{1}_{\|\xi_i\|>\varepsilon\sqrt{n}}|\mathscr{F}_{i-1}\Big]\Big] = \mathbb{E}\big[\|\xi_1\|^2\mathbf{1}_{\|\xi_1\|>\varepsilon\sqrt{n}}\big] \xrightarrow[n\to\infty]{\text{DCT}} 0.$$

We can thus apply a vectorial version of the CLT for martingales difference series obtaining

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\xi_i = \frac{1}{\sqrt{n}}(\sigma^2 C_p)^{-1/2}M_n \xrightarrow[n\to\infty]{d} \mathcal{N}(0, I_p). \tag{8}$$

From (4), (5) we obtain

$$\sigma^{-2}(\boldsymbol{\phi}^* - \boldsymbol{\phi}) = \langle M\rangle_n^{-1}M_n = \langle M\rangle_n^{-1}(n\sigma^2 C_p)(n\sigma^2 C_p)^{-1}M_n.$$

From (5), (8) and Slutsky's lemma we conclude that

$$\sqrt{n}(\boldsymbol{\phi} - \boldsymbol{\phi}^*) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2 C_p^{-1})$$

as desired.

Step 2: we show that $\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*) = o_{\mathbb{P}}(1)$.

By definition,

$$\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*) = \sqrt{n}(\widehat{C}_p^{-1}\widehat{\mathbf{c}}_p - (XX')^{-1}X'Y)$$

and thus

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^*) &= \sqrt{n}\widehat{C}_p^{-1}(\widehat{\mathbf{c}}_p - n^{-1}X'Y) + \sqrt{n}(\widehat{C}_p^{-1} - n(X'X)^{-1})n^{-1}X'Y \\ &= I + II.\end{aligned}$$

We shall prove that both I and II converge to 0 in probability so that, in view of Proposition 1, $\sqrt{n}(\boldsymbol{\phi}^* - \widehat{\boldsymbol{\phi}}) = o_{\mathbb{P}}(1)$. Since $\widehat{C}_p^{-1} \xrightarrow{\mathbb{P}} C_p^{-1}$ (by continuous mapping theorem joined with $\widehat{C}_p \xrightarrow{\mathbb{P}} C_p$), it is enough to prove that $\sqrt{n}(\widehat{\mathbf{c}}_p - n^{-1}X'Y) \xrightarrow{\mathbb{P}} 0$ to deduce that $I \xrightarrow{\mathbb{P}} 0$ (again by Proposition 1). Observe that the i-th component of the vector $\sqrt{n}(\widehat{\mathbf{c}}_p - n^{-1}X'Y)$ is

$$\frac{1}{\sqrt{n}}\Big(\sum_{k=1}^{n-i}X_kX_{k+i} - \sum_{k=1}^{n}X_{k-i}X_k\Big) = \frac{1}{\sqrt{n}}\sum_{k=1}^{i}X_kX_{k+i}.$$

Since $\mathbb{E}[X_k^2] = \mathbb{E}[X_0^2] < \infty$ for any k, $X_k = O_{\mathbb{P}}(1)$ and thus, applying Proposition 1, we deduce that $\sqrt{n}(\widehat{\mathbf{c}}_p - n^{-1}X'Y) = o_{\mathbb{P}}(1)$. Let us now treat the term II. We shall now demonstrate that

$$\sqrt{n}\|\widehat{C}_p^{-1} - n(X'X)^{-1}\| = o_{\mathbb{P}}(1), \tag{9}$$

where, for any matrix A, $\|A\| = \sqrt{\mathrm{Tr}(AA')}$ is the Frobenius norm of A. For that, we begin by noticing that

$$\begin{aligned}\sqrt{n}\|\widehat{C}_p^{-1} - n(X'X)^{-1}\| &= \sqrt{n}\|\widehat{C}_p^{-1}(n^{-1}(X'X) - \widehat{C}_p)n(X'X)^{-1}\| \\ &\le \sqrt{n}\|\widehat{C}_p^{-1}\|\|n^{-1}(X'X) - \widehat{C}_p\|\|n(X'X)^{-1}\|.\end{aligned}$$

Since $\widehat{C}_p^{-1} \xrightarrow{\mathbb{P}} C_p^{-1}$ and $n(X'X)^{-1} \xrightarrow{\mathbb{P}} C_p^{-1}$, it is enough to show that

$$\sqrt{n}\left\|\frac{1}{n}(X'X) - \widehat{C}_p\right\| = o_{\mathbb{P}}(1). \tag{10}$$

Observe that the element in position (i,j) of the $p \times p$-matrix $\frac{1}{n}(X'X) - \widehat{C}_p$ is given by

$$\frac{1}{n}\left(\sum_{k=1}^{n} X_{k-i}X_{k-j} - \sum_{k=1}^{n-|i-j|} X_k X_{k+|i-j|}\right).$$

In the same way as before, we establish (10) using that $X_k = O_{\mathbb{P}}(1)$ for any k. Combining (9) with the fact that $n^{-1}X'Y \xrightarrow{\mathbb{P}} \mathbf{c}_p$, we get that II converges to 0. □

References

[1] Brockwell, P. and Davis, R. Time Series: Theory and Methods, Springer, 2006.